HAL
open science

Une plateforme de développement d'espaces webs sémantiques communautaires dédiés au partage de ressources multimédia

Francky Trichet, Yohan Nizon

To cite this version:

Francky Trichet, Yohan Nizon. Une plateforme de développement d'espaces webs sémantiques communautaires dédiés au partage de ressources multimédia. 18es Journées Francophones d'Ingénierie des Connaissances, Apr 2007, Grenoble, France. pp.245-257. hal-00378319

HAL Id: hal-00378319

https://hal.science/hal-00378319v1

Submitted on 24 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une plateforme de développement d'espaces webs sémantiques communautaires dédiés au partage de ressources multimédia

F. Trichet, Y. Nizon

LINA - Laboratoire d'Informatique de Nantes Atlantique (FRE-CNRS 2729)
Université de Nantes - Equipe Connaissances et Décision (COD)
2, rue de la houssinière - BP 92208 - 44322 Nantes cedex 03
francky.trichet@univ-nantes.fr

Résumé. OSIRIS (*Ontology-based Systems for Semantic Information Retrieval and Indexation dedicated to communautary and open web Spaces*) est un outil dédié au développement d'espaces web communautaires permettant l'annotation et la recherche sémantique de ressources multimédia partagées. Fondé sur l'utilisation d'ontologies denses associées à des thésaurii, OSIRIS permet de décrire sémantiquement une ressource et son contenu en s'appuyant sur un modèle d'annotation intuitif fondé sur le triplet (Sujet, Verbe, Objet), puis de représenter formellement ce contenu à l'aide de graphes conceptuels. Dans ce contexte d'annotation sémantique guidée par des ontologies et thésaurii, OSIRIS permet de prendre en compte plusieurs points de vue sur une même ressource (*i.e.* représentation de plusieurs contenus sémantiques) et de faire cohabiter plusieurs ontologies (couvrant des domaines connexes ou non) au sein d'un même espace. Développé à partir de l'intégration des technologies du Web Sémantique et du Web 2.0, OSIRIS se veut un outil support au développement d'espaces web sémantiques, communautaires et ouverts.

Mots-clés. ontologie, thésaurus, annotation, recherche d'information, graphes conceptuels, webs sémantiques communautaires, multimédia.

1. Introduction

La dimension collective et interactive du Web 2.0 couplée à la légèreté de ces outils a permis l'essor de nombreuses plateformes de partage de ressources multimédia telles que Flickr pour les images ou YouTube et Dailymotion pour les vidéos. Le succès de ces plateformes (en termes de nombre de ressources recensées et nombre d'utilisateurs fédérés) doit cependant être nuancé au regard de la pauvreté des systèmes d'aide à la recherche proposés. En effet, les moteurs mis à disposition des membres de ces communautés d'échanges sont uniquement fondés sur l'utilisation de mots clés (communément appelés *tags* identifiant des folksonomies, *i.e.* des systèmes de classification collaboratifs, décentralisés, libres de toute contraintes et spontanés) directement associés manuellement par les utilisateurs aux ressources partagées. Outre les limites classiques des systèmes par mots clés, en particulier la pauvreté de description sémantique sous-jacente à un ensemble de termes et par conséquent l'impossibilité de mettre en œuvre des processus de recherche d'information sémantique, ces systèmes souffrent d'un manque d'ouverture quant à l'exportation et la pérennité des tags qui restent « propriétaires »

des plateformes et ainsi ne peuvent être exportés lorsque les ressources sont dupliquées.

OSIRIS (*Ontology-based Systems for Semantic Information Retrieval and Indexation dedicated to communautary and open web Spaces*) est un outil qui permet de développer des plateformes à la croisée du Web 2.0 et du Web Sémantique. Dédiée au développement de systèmes d'aide à l'annotation et à la recherche sémantique de ressources multimédia partagées au sein d'espaces web communautaires, cet outil est fondée sur l'utilisation d'ontologies denses [Fürst, Trichet 2006] associées à des thésaurii. Il permet de décrire sémantiquement le contenu d'une ressource multimédia (par exemple, cette photographie de Doisneau représente « *Une femme qui embrasse un homme sur une célèbre place parisienne* »), puis de représenter formellement ce contenu à l'aide de graphes conceptuels [Sowa 1984]. Pour une même ressource, OSIRIS permet également de prendre en compte plusieurs points de vue (*i.e.* représentation de plusieurs contenus sémantiques) qui peuvent être définis en utilisant plusieurs ontologies (caractérisant des domaines connexes ou non). Ainsi, OSIRIS permet, lors du processus d'annotation, d'utiliser plusieurs ontologies qui, lors du processus de recherche, sont utilisées conjointement et de façon transparente grâce à la possibilité de définir des liens d'équivalence entre les concepts et/ou les relations des différentes ontologies considérées (permettant ainsi l'interopérabilité sémantique). Par ailleurs, OSIRIS est fondé sur l'utilisation d'ontologies denses, *i.e.* des ontologies qui en plus d'intégrer les concepts et relations (structurés au sein de hiérarchies fondées sur la relation de Spécialisation/Généralisation) caractérisant le domaine considéré, incluent également les axiomes (règles et contraintes) qui régissent ce dernier. Ceci lui confère la possibilité d'enrichir automatiquement les annotations associées manuellement à une ressource par l'application des axiomes captant les connaissances inférentielles du domaine.

Dans sa version actuelle, OSIRIS permet uniquement de déployer des Webs Sémantiques communautaires dédiés au partage d'images, de sons et de documents bureautiques. Les formats pris en compte sont le JPEG, le MP3 et les formats OpenOffice et Office 2007. Le choix de ces formats est motivé par le fait qu'il est possible de stocker les annotations sémantiques (représentées en termes de graphes conceptuels) au sein même des fichiers via l'utilisation des normes IPTC[1] pour le JPEG, ID3[2] pour les MP3 et ODF[3] pour les documents OpenOffice et Office 2007. Ces normes permettent d'associer des méta-données aux images et sons et de stocker ces dernières au sein même des fichiers. Elles sont utilisées à l'heure actuelle par la plupart des outils de gestion d'images/sons dits « grand public », par exemple Picasa, Winamp ou Xnview. Mais cette utilisation se limite à l'association de mots-clés, ce qui ne permet pas de représenter précisément et formellement le contenu sémantique des ressources. OSIRIS vise à répondre à ce manque en intégrant

[1] http://www.iptc.org
[2] http://www.id3.org
[3] ODF (Open Document Format for Office Applications) - http://www.oasis-open.org/

l'utilisation d'ontologies de domaine dans le processus d'indexation et de recherche. Par ailleurs, conserver les annotations au sein même des fichiers rend notre système beaucoup plus ouvert que les systèmes actuels du web 2.0 où les tags de chaque ressource ne sont utilisables que dans le système considéré.

La suite de cet article est structurée comme suit. La section 2 introduit les fondements de base de nos travaux : les ontologies denses associées à des thésaurii et leurs représentations dans le cadre du modèle des Graphes Conceptuels. La section 3 présente respectivement le modèle d'annotation adopté, le processus d'annotation (manuel et automatique), le processus de recherche et une méthode d'alignement d'ontologies fondée sur une approche extensionnelle. Ces différentes fonctionnalités sont illustrées dans le cadre d'une plateforme dédiée à l'Histoire de l'Art.

2. Contexte des travaux

2.1. Ontologies denses

Actuellement au coeur de nombreuses applications, en particulier le Web Sémantique, les ontologies ont pour objectif de supporter la gestion des connaissances et le raisonnement sur ces connaissances, dans une optique d'interopérabilité sémantique entre agents humains et/ou artificiels. Néanmoins, la plupart des travaux actuels relevant de l'ingénierie des ontologies se limitent à la construction d'ontologies qualifiées de légères (*lightweight ontologies*) car composées simplement d'une hiérarchie de concepts (éventuellement enrichie de propriétés telles que l'exclusion ou l'abstraction) parfois associée à une hiérarchie de relations (éventuellement enrichie de propriétés algébriques). Allégées en sémantique, ces ontologies ne permettent pas de prendre en compte toutes les connaissances d'un domaine donné, en particulier les règles et les contraintes régissant ce dernier et fixant ainsi l'interprétation des concepts et relations le caractérisant. Ce déficit de sémantique (préjudiciable à différents niveaux) s'explique en partie par le faible pouvoir d'expression du langage OWL [OWL 2004]. En effet, ce standard, devenu incontournable pour représenter et partager des ontologies depuis 2004, a indirectement orienté les travaux qui pour la plupart se sont focalisés sur des ontologies légères, en délaissant complètement les connaissances liées à l'inférence (principalement règles et contraintes), tant d'un point de vue représentation de connaissances que d'un point de vue mise en oeuvre de ce type de connaissances dans le cadre de Systèmes à Base de Connaissances.

Dans nos travaux, nous nous intéressons plus particulièrement aux *ontologies denses* sémantiquement parlant (*heavyweight ontologies*), *i.e.* des ontologies qui en plus d'intégrer les concepts et relations (structurés au sein de hiérarchies) du domaine considéré, incluent les axiomes (règles et contraintes) qui régissent ce domaine. L'utilisation conjointe d'une ontologie dense (captant toute la richesse

sémantique d'un domaine via les axiomes) et d'un thésaurus (captant toute la diversité linguistique d'un domaine), caractérisant l'originalité même de la plateforme OSIRIS, s'avère prometteuse dans le cadre d'un système de recherche d'information par mots-clés car l'interprétation du sens (*i.e.* l'identification des concepts et/ou relations mobilisés) sous-jacent à une requête exprimée par un ensemble de termes se retrouve facilitée.

2.2. Modèle des Graphes Conceptuels et langage OCGL

Le modèle des Graphes Conceptuels (GC), introduit par J. Sowa [Sowa 1984], est un modèle de représentation de connaissances, qui appartient à la famille des réseaux sémantiques. Ce modèle est mathématiquement fondé sur la logique et la théorie des graphes [Chein, Mugnier 1992]. Cependant, pour raisonner à l'aide de GC, deux approches peuvent être distinguées : (1) considérer les GC comme une interface graphique pour la logique et donc raisonner à l'aide de la logique et (2) considérer les GC comme un modèle de représentation à part entière disposant de ses propres mécanismes de raisonnement fondés sur la théorie des graphes. Dans le cadre de nos travaux, nous adoptons la seconde approche en utilisant la *projection* (une opération de graphes correspondant à un homomorphisme) comme opérateur de raisonnement ; la projection est complète et cohérente vis-à-vis de la déduction en logique du premier ordre [Chein, Mugnier 1992].

OCGL (*Ontology Conceptual Graphs Language*) [Fürst, Leclère, Trichet 2004] est un langage fondé sur le modèle des GC (une extension du modèle de base) et dédié à la représentation d'ontologies denses. Représenter une ontologie en OCGL consiste principalement à (1) spécifier le vocabulaire conceptuel du domaine considéré et (2) spécifier la sémantique de ce vocabulaire à l'aide d'*Axiomes*. Le vocabulaire conceptuel est composé d'un ensemble de *Concepts* et d'un ensemble de *Relations*. Ces deux ensembles peuvent être structurés soit en utilisant des propriétés conceptuelles bien connues que nous appelons *Schémas d'Axiomes* (couvrant l'expressivité actuelle du langage OWL dans sa version DL comme par exemple les propriétés algébriques des relations, la disjonction de deux concepts, etc.), soit en utilisant des *Axiomes de Domaine* utilisés pour représenter les règles et contraintes.

Les Axiomes de Domaine correspondent aux connaissances inférentielles du domaine qu'il n'est possible de représenter à l'aide des Schémas d'Axiomes, et qui donc ne correspondent pas à des propriétés classiques attestées sur les concepts ou sur les relations. Un Axiome de Domaine est composé d'un graphe *Antécédent* et d'un graphe *Conséquent*, la sémantique formelle d'une telle construction pouvant s'exprimer intuitivement comme suit : « si le graphe Antécédent est attesté vrai, alors le graphe Conséquent est attesté vrai ».

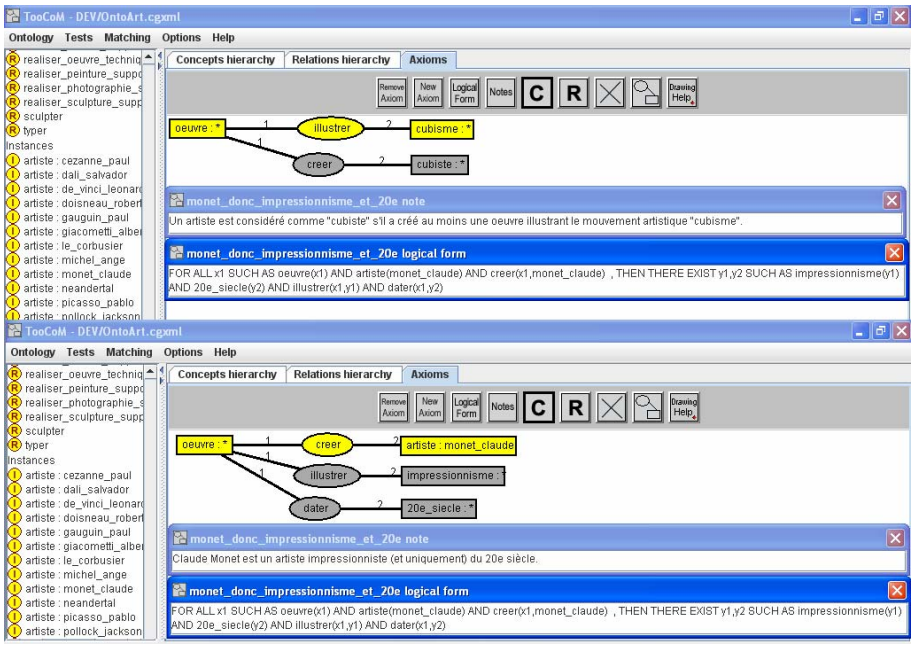

Figure 1. *Exemples d'Axiomes de Domaine représentés en OCGL dans TooCoM.*

La figure 1 présente deux Axiomes de Domaine exprimés en OCGL et permettant respectivement de représenter les connaissances suivantes (inhérentes à une ontologie OntoArt dédiée à l'Histoire de l'Art) : « *Un cubiste est un artiste ayant créé au moins une œuvre illustrant le courant artistique appelé cubisme* » et « *Toutes les œuvres créées par Claude Monet illustrent le courant impressionniste du 20ème siècle* ». Il est opportun de noter que ces deux axiomes ne sont pas du même niveau (et qu'OCGL permet de prendre en compte ces différences de granularité de représentation) : le premier exprime une connaissance générique, le second une connaissance plus spécifique car impliquant une instance du domaine (en l'occurrence Claude Monet).

OCGL a été implémenté au sein de l'outil TooCoM[4] (*A Tool to Operationalize an Ontology with the Conceptual Graph Model)* dédié à l'édition et à l'opérationnalisation d'ontologies denses [Fürst, 2002]. TooCom a été développé à partir de la plate-forme de manipulation de graphes conceptuels CoGITaNT [Genest, Salvat 1998] qui est au cœur d'OSIRIS. Le format standard de stockage d'une ontologie OCGL est le format CGXML[5].

---

[4] TooCoM est disponible sur SourceForge : http://sourceforge.net/projects/toocom/

[5] CGXML est le format XML de représentation des graphes conceptuels intégré à la plateforme Cogitant : http://cogitant.sourceforge.net

3. Plateforme OSIRIS

3.1. Modèle d'annotation adopté

Le modèle d'annotation adopté au sein d'OSIRIS repose sur le triplet {*Sujet / Verbe / Objet*} permettant de représenter le contenu de phrases simples exprimées en langage naturel telles que « *Un homme qui embrasse une femme* ». Au sein de ce triplet, *Sujet* et *Objet* correspondent à des concepts et *Verbe* une relation de l'ontologie considérée (cf. figure 3). Chaque ressource peut ainsi être décrite sémantiquement par un ensemble de triplets, où chaque triplet est défini par rapport à une ontologie particulière, sachant que plusieurs ontologies (pour un même domaine ou des domaines connexes) peuvent cohabiter au sein d'une même plateforme développée à l'aide d'OSIRIS. Il est également possible d'associer chaque annotation à un membre de la communauté.

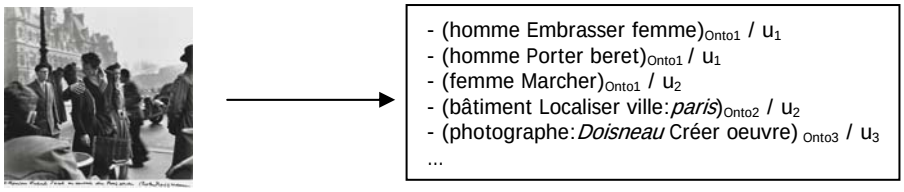

Figure 2. *Application du modèle d'annotation au célèbre « Baiser de l'hôtel de ville » de Doisneau.*

La figure 2 illustre l'application de ce modèle d'annotation au célèbre « *Baiser de l'hôtel de ville* » de Doisneau. Dans cet exemple, le premier utilisateur u_1 annote la photographie en utilisant l'ontologie $Onto_1$ et décrit « *Un homme qui embrasse une femme* » (homme étant le concept *Sujet*, Embrasser la relation *Verbe* et femme le concept *Objet*) et « *Un homme qui porte un béret* ». Le second utilisateur u_2 annote en utilisant deux ontologies $Onto_1$ et $Onto_2$, et décrit « *Une femme qui marche* », sans préciser explicitement où (*i.e.* absence de concept *Objet*), et « *Un bâtiment situé dans une ville nommée Paris* » (*i.e.* une précision sur le concept *Objet* ville qui ici est instancié par *Paris*). Enfin, le dernier utilisateur u_3 annote non pas le contenu de la photographie mais la photographie en tant que telle en précisant que celle-ci est une œuvre créée par le photographe *Doisneau* (utilisation similaire de l'instanciation du concept *Sujet* pour préciser l'annotation, en l'occurrence photographe:*Doisneau*). Ce dernier point illustre par ailleurs que notre modèle peut être utilisé à la fois pour décrire le contenu d'une ressource et pour décrire la ressource en tant que telle, ce qui le rend flexible et ouvert.

Ce modèle d'annotation, intuitif, facilement compréhensible et possédant une correspondance forte avec le modèle des Graphes Conceptuels, permet aux

utilisateurs de décrire le contenu d'une ressource sous plusieurs angles, et éventuellement en utilisant plusieurs ontologies du même domaine (et dans ce cas pouvant être plus généralistes, plus spécialisées et/ou du même niveau de description mais développées par des communautés différentes), de domaines connexes et/ou non connexes. Ceci permet de mettre en œuvre un processus d'annotation à la fois multi-utilisateurs, multi-points de vue et multi-ontologies.

3.2. Annotation de ressources

3.2.1 Approche manuelle

L'annotation d'une ressource importée par un utilisateur au sein d'OSIRIS débute par la sélection d'une ontologie préalablement importée par l'administrateur[6]. Ce processus d'annotation consiste à identifier des triplets (*Sujet,Verbe,Objet*) où *Sujet* et *Objet* correspondent à des concepts et *Verbe* une relation de l'ontologie considérée O. Pour ce faire, l'utilisateur peut soit (1) directement naviguer au sein des hiérarchies de concepts et de relations de O, soit (2) exprimer librement une liste de termes qui seront alors comparés aux entrées du thésaurus associé à O. Dans le premier cas, l'utilisateur est guidée par l'interface car dès lors où il a identifié le concept correspondant au *Sujet* ou *Objet* de son annotation, alors seules les relations possédant en signature le concept sélectionné (ou des concepts plus spécifiques) lui sont proposées. Il en est de même lorsque l'utilisateur commence par identifier la relation associée au *Verbe* de son annotation : seuls les concepts compatibles à la signature lui sont alors accessibles. Dans le second cas, qui vise à offrir plus de liberté et « d'ouverture linguistique » quant à l'expression du contenu, OSIRIS utilise les thésaurii pour cibler les concepts et relations sous-jacents, puis propose les éventuelles annotations (*Sujet,Verbe,Objet*) correspondantes (s'il en existe). L'utilisateur sélectionne alors la (ou les) annotation(s) la (les) plus pertinente(s) par rapport à son objectif de recherche.

La figure 3 illustre ce processus d'annotation dans le cadre du *« Baiser de l'hôtel de ville »* de Doisneau. Les annotations déjà définies par l'utilisateur sont précisées dans l'onglet « Contenu » (« *Un homme qui embrasse une femme* », « *Un homme qui porte un béret* » et « *Une photographie de l'artiste Robert_Doisneau* »). L'annotation en cours (onglet « Annoter vos fichiers ») est « *Un monument daté du 14ème siècle* ». Il est important de souligner qu'il est également possible de lier les annotations entre-elles en explicitant, par exemple, que deux instances de concepts impliqués dans deux annotations différentes sont identiques ou différents. Ceci

[6] OSIRIS permet d'importer des ontologies aux formats OWL ou CGXML, sachant que seul le format CGXML lié au langage OCGL permet de prendre en compte les axiomes et donc de mettre en œuvre l'enrichissement automatique des annotations. Il permet également d'importer des thesaurii au format SKOS, un format de représentation RDF dédié aux ressources de type thésaurus - http://www.w3.org/2004/02/skos/.

permet, dans le cadre de l'exemple de la figure 2, de préciser que l'homme qui porte le béret est différent de celui qui embrasse la femme.

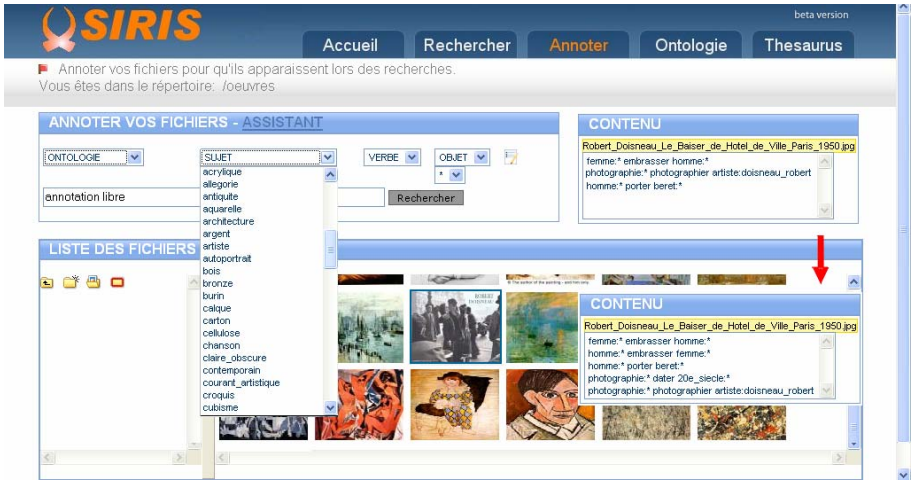

Figure 3. *Illustration du processus d'annotation.*

Chaque annotation est enregistrée (au format XML) au sein même des fichiers via les normes IPTC, ID3 et ODF. OSIRIS permet également d'exporter automatiquement les mots-clés au format IPTC, ID3 et ODF via la comparaison des annotations (*Sujet,Verbe,Objet*) aux entrées des thésaurii.

3.2.2 Approche automatique

Les ontologies denses manipulées par OSIRIS incluent par essence des axiomes. L'application de ces axiomes sur les annotations préalablement définies « à la main » permet de mettre en œuvre un processus d'enrichissement automatique des annotations des ressources. La figure 3 (onglet « Contenu » ajouté manuellement sur la figure) illustre les résultats obtenus suite à l'application automatique des axiomes des deux ontologies cohabitant au sein de cette plateforme : OntoArt dédiée à l'Histoire de l'Art et incluant la connaissance « *Toute œuvre de Doisneau est datée du 20ème siècle* » et OntoCourant, une ontologie recouvrant les phénomènes de la vie courante et intégrant par exemple la relation Embrasser définie entre deux concepts Etre_Humain (se spécialisant en Homme et Femme). L'application de la symétrie de la relation Embrasser (représentée par un Schéma d'Axiome en OCGL) produit la nouvelle annotation « *Femme Embrasser Homme* » et l'application de la connaissance de OntoArt (représentée par un Axiome de Domaine) produit l'annotation « *Photographie Dater 20_Siecle* ».

Par ailleurs, lors de l'import d'une ressource par un utilisateur, OSIRIS vérifie si celle-ci comporte déjà des mots clés (qui peuvent lui avoir été associés à partir

d'autres plateformes) via l'utilisation des normes IPTC, ID3 ou ODF. Si tel est le cas, OSIRIS effectue une analyse de ceux-ci pour tenter de retrouver automatiquement des annotations en s'appuyant sur les entrées des thésaurii. Le résultat de cette analyse conduit à la proposition d'un ensemble d'annotations potentielles que l'utilisateur doit valider.

3.3. Recherche de ressources

La recherche de ressources débute par la formulation d'une requête (*Sujet,Verbe,Objet*), ou d'un ensemble de requêtes connectées entre elles par les opérateurs ET et OU. Pour formuler ses requêtes, l'utilisateur peut soit naviguer au sein des hiérarchies, soit exprimer librement des termes qui seront ensuite comparés aux entrées des thésaurii pour retrouver les concepts et relations sous-jacents. Notons qu'il est possible d'exprimer des requêtes dites partielles (*i.e.* fondées partiellement sur le triplet) et respectant les formats suivants : (*Sujet*) ou (*Objet*), (*Sujet,Verbe*) ou (*Verbe,Objet*).

A chaque requête C correspond un (ou plusieurs) graphe(s) conceptuel(s). La recherche de ressources respectant les critères définis par la requête est effectuée en utilisant l'opérateur de projection des GC : une ressource R_i satisfait une requête C s'il existe (au minimum) une projection du graphe conceptuel représentant C sur un des graphes représentant les annotations associées à R_i.

Figure 4. *Illustration du processus de recherche : « Les œuvres contemporaines représentant des femmes ». Les œuvres retrouvées sont des photographies, des sculptures ou des peintures.*

La figure 4 illustre un exemple de scénario de requêtes où les critères sont (1) « U*ne œuvre contemporaine* » représenté par le graphe conceptuel (Œuvre:* Dater Contemporain:*), où Œuvre et Contemporain sont des concepts et Dater une relation ET (2) « *dont le contenu incarne une femme* » représentée par le graphe conceptuel composé uniquement du concept (Femme:*). Toute ressource R pour laquelle il existe une projection de chacun de ces deux graphes dans (au moins) un des graphes représentant une des annotations de R est considérée comme pertinente pour la requête.

OSIRIS permet également d'effectuer des recherches sur des instances de concepts. Par exemple, la requête «*Quelles sont les peintures créées par l'artiste Picasso* ?» (se traduisant par « peinture Créer artiste:*Picasso* ») permet, par le fait d'avoir préciser le concept peinture et non le concept oeuvre, de spécifier que la recherche doit uniquement se focaliser sur les oeuvres de Picasso en tant que peintures, et non sur ses autres œuvres comme les sculptures. OSIRIS permet également d'exprimer des requêtes identifiant des relations dénuées de concepts (par exemple, « Embrasser »), requêtes que nous qualifions de partielles.

3.4. Alignement d'ontologies

L'objectif de l'alignement d'ontologies est de découvrir et d'évaluer des liens conceptuels (par exemple des identités, des subsomptions, ou des disjonctions) entre primitives conceptuelles (*i.e.* concepts et relations) de deux ontologies supposées bâties sur des domaines connexes [Shvaiko 2005].

Etant donné qu'il est possible de faire cohabiter plusieurs ontologies au sein d'une même plateforme et qu'il s'avère nécessaire de rendre transparent et efficient l'utilisation même des ces multiples ontologies lors des phases de recherche d'information, OSIRIS offre une fonctionnalité (dédiée à l'administrateur) permettant de répondre à cette problématique d'alignement en calculant automatiquement un ensemble d'appariements de concepts et de relations (issus d'ontologies différentes) pondérés par un coefficient de vraisemblance. Notre approche, qualifiée d'extensionnelle (*i.e.* comparaison des extensions des concepts et relations des ontologies) repose sur le principe suivant : lorsqu'une même ressource R est annotée en utilisant plusieurs ontologies[7], alors il s'avère pertinent d'apparier les concepts et relations des ontologies impliquées dans les annotations de R, étant donné que ces annotations multi-ontologies visent à représenter le même contenu « sémantiquement parlant ».

Notre algorithme prend en entrée plusieurs ontologies (préalablement importées sous OSIRIS) et un ensemble de ressources annotées à l'aide de ces dernières et

7 Ou lorsqu'un ensemble de ressources sont sélectionnées graphiquement par l'utilisateur et qualifiées de « sémantiquement proches en terme de contenus », ce qui permet à notre algorithme d'alignement de rester opérationnel même lorsque les ressources ne sont pas, de façon unitaire, annotées à l'aide de plusieurs ontologies.

produit en sortie, pour chaque combinaison d'ontologies possibles O_i et O_j, deux ensembles d'appariements respectivement dédiés aux concepts et relations : $\{(C_i,C_j,VC_k)\}$ et $\{(R_i,R_j,VR_k)\}$ où C_i et C_j (resp. R_i et R_j) sont des concepts (resp. des relations) respectifs de O_i et O_j et VC_k (resp. VR_k) le coefficient de vraisemblance entre C_i et C_j (resp. R_i et R_j). Le calcul des coefficients de vraisemblance repose sur le principe suivant : si deux annotations $(S1,V1,O1)_{Onto1}$ et $(S2,V2,O2)_{Onto2}$, *i.e.* des triplets partiels ou complets de (*concept relation concept*) provenant de deux ontologies différentes $Onto_1$ et $Onto_2$, sont associées à une même ressource (ou à deux ressources considérées comme proches en terme de contenus par l'utilisateur) alors incrémenter les coefficients de 1 des appariements (S1,S2), (V1,V2) et (O1,O2) si ceux-ci existent déjà, sinon créer ces appariements et initialiser les coefficients à 1.

Appliqué à un ensemble volumineux de ressources, ce principe conduit à l'identification d'un grand nombre d'appariements. En utilisant un seuil de sélection pour conserver uniquement les appariements à forts coefficients, cette approche de l'alignement extensionnelle conduit à l'identification automatique d'équivalences entre les concepts et relations de deux ontologies différentes, équivalences qui doivent être validées par l'administrateur de la plateforme pour devenir effectives lors des processus d'annotation et de recherche. Notons que ce dernier a également la possibilité, via un module d'administration des ontologies, de définir directement « à la main » ces équivalences, ce qui permet d'étendre les capacités de recherche de la plateforme, tout en restant complètement transparent pour les utilisateurs.

4. Conclusion

OSIRIS est un outil support à l'émergence d'espaces web de partage de ressources multimédia (limitées à l'heure actuelle aux formats JPEG, MP3 et documents OpenOffice et Office 2007). Ces espaces se veulent à la fois (1) sémantiques, car reposant sur une représentation formelle des domaines de connaissances considérés, (2) communautaires, car visant à fédérer des utilisateurs partageant des définitions consensuelles des domaines impliqués (capturées aux sein d'ontologies) et des pratiques communes sur ces domaines et (3) ouverts tant sur la dimension technique (possibilité d'exporter les annotations via le stockage de ces dernières au sein même des fichiers) que sur la dimension conceptuelle (possibilité de faire cohabiter plusieurs ontologies).

L'annotation sémantique fondée sur l'utilisation des graphes conceptuels n'est pas une approche novatrice. En effet, plusieurs travaux s'inscrivent dans ce mouvement. Citons à titre d'exemple [Bocconi & al. 2005], [Crampes, Ranwez. 2000] ou [Issac, 2004]. L'originalité de nos travaux ne se situe donc pas l'approche adoptée, mais dans le contexte dans lequel cette approche est envisagée. En effet, contrairement aux travaux similaires, OSIRIS offre la possibilité de faire cohabiter

et évoluer plusieurs ontologies (connexes ou non) au sein d'un même espace et ces dernières peuvent être affinées par les membres de la communauté.

Nos travaux se poursuivent actuellement sur l'étude plus approfondie des *tags* associés aux ressources au sein des plateformes Web 2.0 pour, outre enrichir automatiquement les thésaurii et affiner notre approche d'aide à l'annotation, découvrir d'éventuels manques dans les ontologies considérées, et ce en analysant les traces d'usages de la plateforme. Notre hypothèse est que les tags définis et partagés par les communautés sont profondément porteurs de sens et donc vecteurs des évolutions des domaines de connaissances sous-jacents. Dans ce cadre, il nous paraît pertinent de s'appuyer sur ce type de matériau pour prendre en compte la problématique de l'évolution des ontologies, véritable clé de voûte de la popularisation des plateformes sémantiques et participatives telle que OSIRIS.

Bibliographie

BOCCONi S., NACK F. and HARDMAN L. (2005). « Supporting the Generation of Argument Structure within Video Sequence », Proceedings of the *Sixteenth ACM Conference on Hypertext and Hypermedia*, pp 75-84.

CHEIN, M. AND MUGNIER, M.L. (1992). « Conceptual Graphs: fundamental notions », *Revue d'Intelligence Artificielle (RIA)*, vol 6(4), p. 365-406, Hermès.

CRAMPES M. and RANWEZ S. (2000). « Ontology-Supported and Ontology-Driven Conceptual Navigation on the World Wide Web », Proceedings of the *eleventh ACM on Hypertext and hypermedia*, pp 191–199.

FÜRST F., TRICHET F. (2006). « Heavyweight Ontology Engineering », *Proceedings of the 5th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE'2006). LNCS 4277*, pp. 38–39. Springer-Verlag (Berlin Heidelberg).

FÜRST F., LECLERE M, TRICHET F. (2004). « Operationalizing domain ontologies: a method and a tool », *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'2004)*, pp 318-322. IOS Press.

FÜRST F. (2002). « Contribution à l'Ingénierie des Ontologies : une méthode et un outil d'opérationnalisation », *Thèse de l'Université de Nantes*.

GENEST D., SALVAT E. (1998). « A Platform allowing typed nested graphs: how CoGITo became CoGITaNT », *Proceedings of the 6th International Conference on Conceptual Structures. LNAI 1453*, pp 154-161. Springer-Verlag.

ISAAC A. (2004) « Conception et utilisation d'ontologies pour l'indexation de documents audiovisuels ». Thèse Université de la Sorbonne Paris IV.

SHVAIKO P., EUZENAT J. (2005). « A Survey of Schema-based Matching Approaches », Journal on Data Semantics IV. LNCS 3730, p146-171. Springer.

SOWA J. (1984). *Conceptual Structures: information processing in mind and machine*, Handbook, Addison-Wesley.